Predicting natural product value, an exploration of anti-TB drug space

Journal:	*Natural Product Reports*
Manuscript ID:	NP-REV-02-2014-000021.R1
Article Type:	Highlight
Date Submitted by the Author:	26-Mar-2014
Complete List of Authors:	Dashti, Yousef; Griffith University, Eskitis Institute for Drug Discovery Grkovic, Tanja; Griffith University, Eskitis Institute for Drug Discovery Quinn, Ron; Griffith University, Eskitis Institute

Journal Name

RSCPublishing

HIGHLIGHT

Predicting natural product value, an exploration of anti-TB drug space

Cite this: DOI: 10.1039/x0xx00000x

Yousef Dashti,[a] Tanja Grkovic[a] and Ronald J Quinn[a*]

Received ooth January 2012,
Accepted ooth January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

Mycobacterium tuberculosis (Mtb) still remains a deadly pathogen two decades after the announcement of tuberculosis (TB) as a global health emergency by the World Health Organization. In last few years new drug combinations are showing promising potential to significantly shorten TB treatment times. However there are very few new chemical entities being developed to treat this global threat. From January 1990 to December 2012, 949 anti-mycobacterium natural products were reported in the literature. Here we present a perspective based on an analysis of the drug-like properties of the reported anti-mycobacterium natural products in order to assess drug potential.

1 Introduction

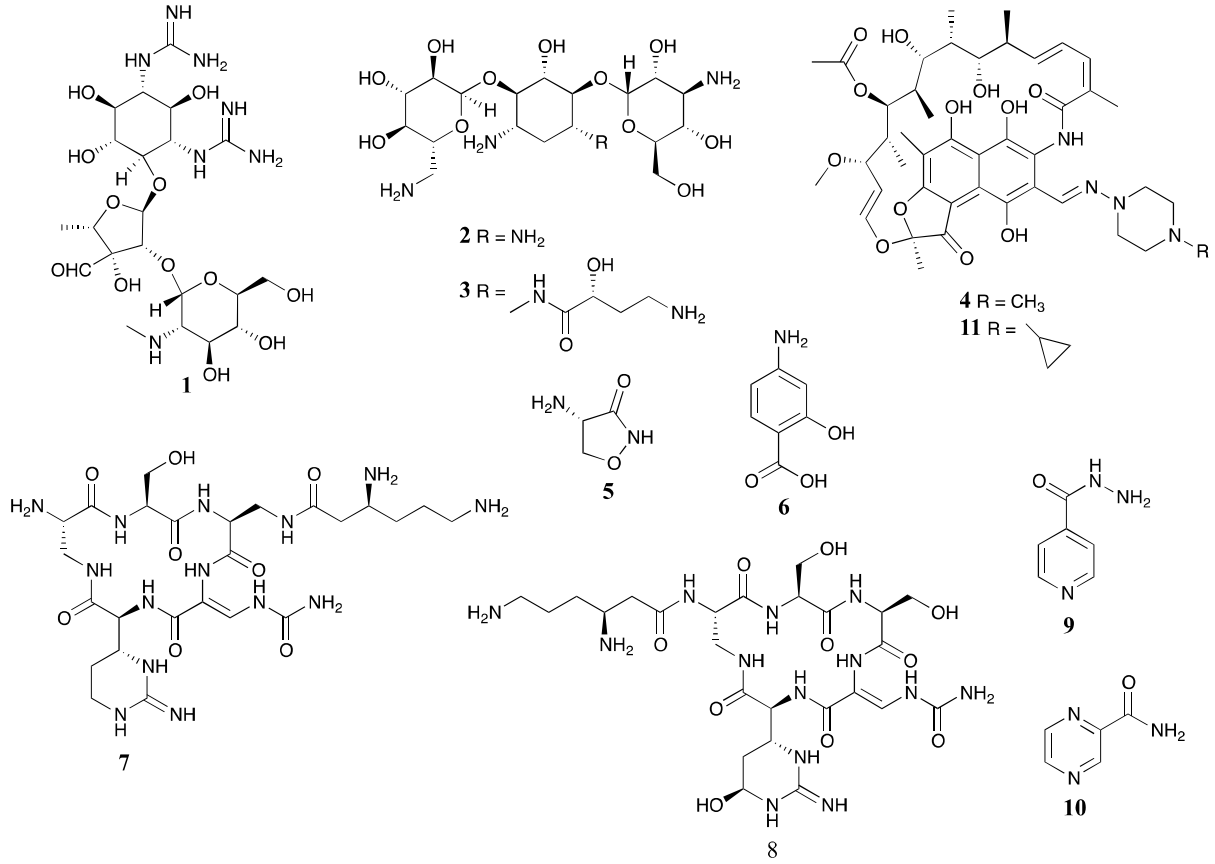
Tuberculosis (TB), an illness that most commonly results from infection by *Mycobacterium tuberculosis* (Mtb), has been responsible for the death of almost 30 million people around the world in the last decade.[1] Currently Mtb is the second most deadly pathogen after human immunodeficiency virus (HIV). One third of the world's population is thought to be infected with Mtb and 10% of those carry a lifetime risk of developing the disease.[1]

TB can be cured by long multidrug regimes (6-12 months) but, due to a number of factors including limited access to diagnosis and treatment in developing countries, the spread of HIV, and the emergence of multi- and extensively drug-resistant Mtb strains, tuberculosis still remains a leading cause of death in the world.[2, 3] In addition to the resistance-mediated decline in efficacy, current anti-TB drugs have side effects, difficulties with dosing regimens and they are unable to combat latent TB. New efficacious and affordable chemotherapeutics are needed in order to improve patient outcomes, reduce side effects, shorten treatment regiment, reduce the emergence of resistance strains as well as complications associated with HIV co-infections.

Nature is a generous source of active compounds. Investigation of the correlation between biosynthetic enzymes and therapeutic targets identified that an imprint of the recognition of biosynthetic protein surfaces is transferred to recognition of therapeutic protein target surfaces. A similar protein fold topology has been demonstrated in a study of different targets of the same natural product.[4-6] This correlation of biosynthetic and target proteins explains both the success of compound libraries based on natural product starting points and why natural products are validated starting points for drug design.[7-9] Compared to synthetic compounds, nature-derived molecules have a higher degree of diversity, and an average increase of heteroatoms. A study by Feher and Schmidt revealed that natural products are more similar to drugs then molecules obtained from combinatorial synthesis.[10] The majority of natural products are biologically active and have better ADME properties (adsorption, distribution, metabolism, and excretion) compared to combinatorial synthetic compounds.[10] However, as concluded in an investigation by Shoichet *et al*., 83% of core ring scaffolds present in natural products (NPs) are absent from the majority of screening libraries and consequently the large compound libraries used in high throughput screening (HTS) may not reflect the rich diversity of smaller natural product-based libraries.[11] It has been suggested that inclusion of molecules containing scaffolds present in NPs in these compound libraries would improve hit rates.[11]

The importance of natural products in antibacterial drug discovery is undeniable. As documented by Newman and Cragg,[12] just 3 of the 20 antibacterial classes used in the clinic are from synthetic sources and about 70% of antibiotics are either natural products or semi-synthetic natural product derivatives. The peptide actinomycin, a streptomycete-derived compound, was the first reported natural product that inhibited *in vitro* growth of Mtb.[13] The aminoglycoside streptomycin **1**, isolated from *Streptomyces griseus*, was the first successful

antibiotic used in chemotherapy of tuberculosis.[14] Kanamycin (**2**), a compound related to streptomycin,[15] and its semi-synthetic analogue amikacin (**3**)[16] are part of the second line chemotherapy of TB. Rifampicin (**4**), a semi-synthetic analogue



of rifamycin,[17] cycloserine (**5**),[18] *p*-aminosalisylic acid (**6**),[19] and cyclic peptides capreomycin 1A (**7**)[20, 21] and viomycin (**8**),[22, 23] are other examples of nature-derived anti-tuberculosis drugs. Reports of anti-TB activity of the natural product nicotinamide led to a reinvestigation of the related synthetic compound isoniazid (**9**) which is now one of the most important first line drugs used in tuberculosis chemotherapy. Subsequently an isostere of isoniazid, pyrazinamide (**10**), was synthesized as a potent anti-TB drug.[24]

Pyrazinamide, rifampicin and isoniazid are three out of the four components of first line TB treatment. The aminoglycosides streptomycin, kanamycin and amikacin as well as cyclic peptides, capreomycin and viomycin, are among the second line drugs. These molecules also play an important role in new treatment regimens for improving efficacy and shortening TB chemotherapy.[25] Since these nature-derived compounds in combination with synthetic TB drugs were successful in treatment of TB and have led to an initial decline of TB, investigations for new structural class of antibiotics as anti-tuberculosis agents almost halted in the 1960s after the discovery of rifampicin.[26]

Following the emergence of multi- and extensively-drug resistant Mtb strains in addition to comorbidity with HIV, the number of TB-related deaths has significantly increased. In 1993 the World Health Organization (WHO) declared TB a global health emergency and after two decades the report of WHO in 2012 indicates that TB still remains a global threat.[27, 28]

During the last five years some promising compound candidates entered the TB drug pipeline.[25] Except for a few compounds that were first discovered as anti-mycobacterial molecules and consequently developed as TB drug candidates, many of these are repurposed compounds that were already used for treatment of other diseases. While the majority of approved drugs currently used in TB treatment are nature-derived molecules, all of the TB drug candidates in clinical trials, with the exception of rifapentine (**11**), are from synthetic sources. Many natural products have been reported to have activity against mycobacteria during the last two decades and are covered by a number of recent comprehensive reviews of anti-mycobacterium natural products.[29-34] However, none of these compounds had been developed for treatment of TB. Here, we review the literature and present a perspective based on an analysis of the drug-like properties of the anti-mycobacterium natural products.

2 Physicochemical properties of TB drugs and anti-mycobacterium natural products

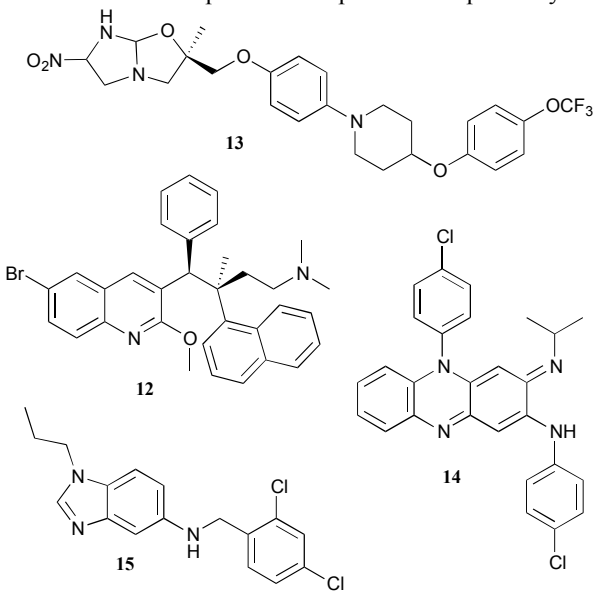
The activity of any molecule is generally dependent upon the compound being able to permeate the cell membrane in order to reach and modify the target. It is possible to predict the cell permeability and bioavailability of each molecule by examining the physicochemical properties of the compound. Lipinski and coworkers analyzed around 90% of orally active drug candidates that reached phase II clinical trials in order to understand which factors are responsible for compound attrition in clinical development. Their investigation resulted in the "rule of five" (Ro5), based on a simple set of easily calculated

physicochemical properties. By comparison of drug-like and non-drug-like compounds, they concluded that for a compound to be drug-like it should have the cut-off numbers of five or multiples of five in four important physicochemical properties. These properties are hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), molecular weight (MW), and the logarithm of the partition coefficient between water and 1-octanol (log *P*). According to Lipinski's rule of five, a drug like molecule should have a molecular weight of 500 Da or less, equal or less than 5 HBDs, equal or less than 10 HBAs, and a partition coefficient, log *P,* of 5 or less. If two or more properties are violated for a compound, there is a high probability of lack of oral activity and bioavailability.[35] However, it is important to note that if a compound passes the Ro5, there is no guarantee that the molecule is drugable. The aim of this rule was to provide a guide for medicinal chemists for better selection and design of compounds, in order to reduce the attrition during clinical development due to unsatisfactory pharmacokinetics.[36] Subsequently, some extensions including rotatable bond count (ROTB) and polar surface area (PSA) or total hydrogen bond count were introduced to complement the Ro5. Rotatable bond count is now one of the widely used properties to predict oral bioavailability and activity. Generally, ligand affinity decreases 0.5 kcal on average for each two rotatable bonds.[37] If a rigid and a flexible ligand can form the same pattern of interaction with a protein (based on hydrogen and hydrophobic interaction), the rigid ligand will show much stronger binding due to lower entropic losses.[38] A molecule with 10 or fewer rotatable bonds, polar surface area equal to or less than 140 Å (hydrogen bond donors and acceptors equal or fewer than 12), will have a high possibility of being a good orally bioavailable drug.[39]

Herein, the four individual Lipinski properties HBD, HBA, MW, and log *P* along with ROTB and PSA were analyzed for 39 TB drugs (approved and candidates in clinical trials) as well as 949 anti-tuberculosis NPs reported between January 1990 to December 2012. InstantJChem 3.0.4. [Instant JChem 3.0.4, 2009 ChemAxon Ltd. (http://www.chemaxon.com)] was used to calculate the physicochemical parameters of each molecule. The following paragraphs are devoted to discussion of the physicochemical properties of TB drugs, followed by the physicochemical properties of anti-mycobacterial NPs and comparison between these two sets.

2.1 Physicochemical properties of TB drugs

The calculated data for physicochemical properties of approved TB drugs and candidates in clinical trials are shown in Table 1. The percentage of TB drugs compliant with Lipinski's rule of five is depicted in Figure 1. Almost 77% of TB drugs comply with all of the Lipinski's parameters or have just one violation while 23 % (nine drugs) have two or more violations. Five of the nine TB drugs that violate the Ro5 are intravenous/injectable drugs (streptomycin, kanamycin, amikacin, capreomycin, and viomycin). The remaining four are orally bioavailable drugs (rifampicin and its analogue refapentine, the synthetic drugs bedaquiline (TMC207) (**12**) and the nitroimidazole delamanid (OPC67683) (**13**). With the exception of the approved drug rifampicin, the other three compounds are drug candidates in clinical trials. Refapentine is a repurposed compound, which is currently in the phase II clinical trials, while bedaquiline and delamanid are new chemical entities in phase II and phase III respectively.[25]



The histograms for molecular weight (MW), calculated log *P*, hydrogen bond donors (HBD) and acceptors (HBA) are depicted in Figure 2 (purple bars). About 51% of TB drugs are distributed between molecular weights of 300-500 Da and are all synthetic molecules. Nature-derived TB drugs either possess the lowest (cycloserine, pyrazinamide, isoniazid, and p-aminosalicylic acid) or the highest molecular weight (rifampicin, rifapentine, capreomycin, and viomycin).

The distribution of the calculated log *P* (Figures. 2b and S1) showed a wide range of variability. The majority of compounds are distributed between log *P* of -3 to 5 and showed a bimodal distribution with peaks in the range of -1 to 1 and 3 to 5. Cyclic peptides, capreomycin and viomycin are the most polar TB drugs with log *P* of around -11. On the other hand, 10% of the TB drugs which all are synthetic compounds have calculated log *P* of more than 5. These include clofazimine (**14**) and the drug candidates TMC207, OPC-67683, C215 (**15**). Overall synthetically-derived TB drugs possess higher Log *P* values compared to the nature-derived TB drugs (Figure S1).

The distribution of hydrogen bond donors (HBD) revealed a maximum at 1 and then steadily decreased, while a wide range of variability was observed in the distribution histogram of hydrogen bond acceptors (HBA) of the TB drugs. Almost 18% percent of the compounds had violations from the HBD and HBA cutoff numbers 5 and 10, respectively. All of these are nature-derived drugs including streptomycin, kanamycin, amikacin, capreomycin and viomycin as well as the rifamycins analogues, rifampicin and rifapentine.

Table 1 Physicochemical properties of 39 approved TB drugs and candidates in clinical trials.

Nature-derived TB drugs | First line drugs | Third line drugs
Synthetic TB drugs | Second line drugs | Candidates in clinical trials

Name	MW	Log *P*	HBD	HBA	PSA	ROTB	Ro5	No. Violations
Rifampicin	822.94	2.77	6	14	220.15	5	Fail	3
Isoniazid	137.14	-0.69	2	3	68.01	1	Pass	0
Pyrazinamide	123.11	-1.23	1	3	68.87	1	Pass	0
Ethambutol	204.31	-0.06	4	4	64.52	9	Pass	0
Streptomycin	581.57	-7.65	14	19	331.43	9	Fail	3
Kanamycin	484.50	-7.06	11	15	282.61	6	Fail	2
Amikacin	585.60	-8.58	13	17	331.94	10	Fail	3
Cycloserine	102.09	-2.42	2	3	64.35	0	Pass	0
p-aminosalicylic acid	153.14	0.83	3	4	83.55	1	Pass	0
Capreomycin	668.71	-11.00	15	14	375.92	10	Fail	3
Viomycin	685.69	-11.05	16	15	390.36	10	Fail	3
Ethionamide	166.24	1.33	1	1	38.91	2	Pass	0
Terizidone	302.29	-0.38	2	6	101.38	4	Pass	0
Prothionamide	180.27	1.77	1	1	38.91	3	Pass	0
Thioacetazone	236.29	0.90	3	2	79.51	3	Pass	0
Ciprofloxacin	331.34	-0.81	2	6	72.88	3	Pass	0
Ofloxacin	361.37	0.65	1	7	73.32	2	Pass	0
Levofloxacin	361.37	0.65	1	7	73.32	2	Pass	0
Clavulanate	199.16	-1.52	2	5	87.07	2	Pass	0
Meropenem	383.46	-4.35	3	6	110.18	5	Pass	0
Clofazimine	472.41	7.81	1	3	27.63	4	Pass	1
Moxifloxacin	401.43	-0.50	2	7	82.11	4	Pass	0
Gatifloxacin	375.39	-0.58	2	7	82.11	4	Pass	0
Linezolid	337.35	0.64	1	5	71.11	4	Pass	0
Sutezolid	353.41	1.22	1	4	61.88	4	Pass	0
AZD5847	465.40	0.78	2	7	125.57	7	Pass	0
SQ109	330.55	4.64	2	2	24.06	9	Pass	0
AU1235	324.34	3.70	2	1	41.13	2	Pass	0
TMC207	555.51	7.13	1	4	45.59	8	Fail	2
DNB1	361.31	2.41	1	7	139.20	8	Pass	0
BTZ043	431.39	3.65	0	7	96.95	3	Pass	0
PA-824	359.26	4.14	0	6	91.33	6	Pass	0
BM212	414.37	4.75	0	2	11.41	4	Pass	0
OPC-67683	534.48	6.14	0	8	103.80	9	Fail	2
Metronidazole	171.15	-0.46	1	4	83.87	3	Pass	0
C215	333.26	6.11	0	1	17.82	5	Pass	1
VI-9376	344.16	4.17	0	4	71.60	2	Pass	0
Rifapentine	877.03	3.56	6	14	220.15	6	Fail	3
377790	260.29	3.60	0	4	76.53	4	Pass	0

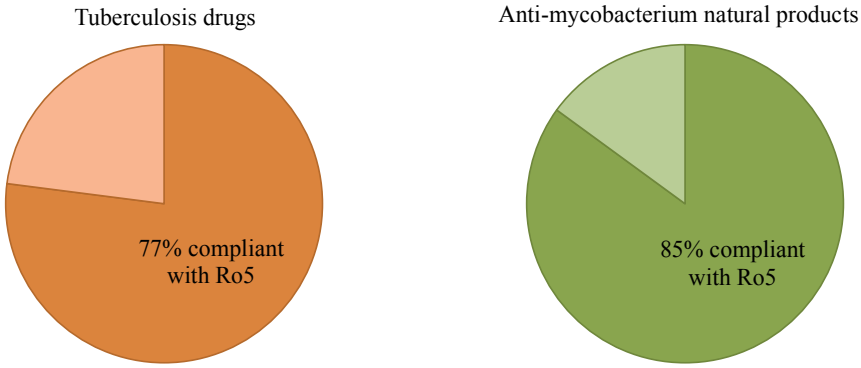


Figure 1 Pie chart presentation of the percentage of TB drugs (left) and anti-mycobacterium natural products (right) compliant with Lipinski's rule of five.

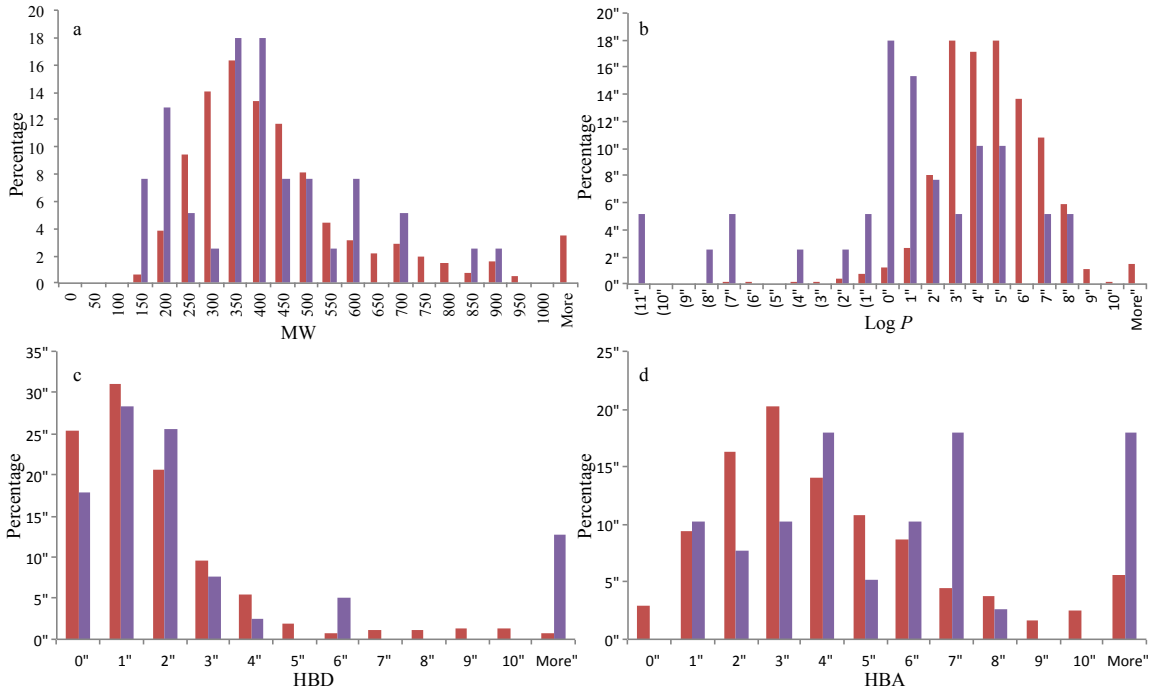


Figure 2 Physicochemical property histograms of TB drugs (purple) and anti-mycobacterial natural products (red): a) molecular weight (MW), b) calculated log *P*, c) hydrogen bond donors (HBD), d) hydrogen bond acceptors (HBA).

It is a common practice to analyze physicochemical properties prior to making chemical libraries. Antibacterial drugs however, due to their higher molecular weight and polarity, have always been considered as an exception to the rule of five.[35, 40] The different components of the human cell wall and the bacteria suggests a requirement for different physicochemical properties of drug compounds for penetration. In a study by O'Shea and Moser the authors concluded that antibacterial compounds, especially for those active against Gram-negative bacteria, have average molecular weights and polarity higher than non-antibacterial compounds.[40] Gram-positive and -negative bacteria have different cell wall architecture and consequently molecules with ability to penetrate these two types of bacteria have different properties and occupy different physicochemical space.[40] On the other hand, although *Mycobacterium tuberculosis,* based on its conserved gene content, is more related to Gram-negative bacteria its unique membrane architecture has characteristics of both Gram-positive and -negative bacteria.[41] This may be the reason that marketed anti-TB drugs are widely distributed within physicochemical space and do not fall into any defined physicochemical area. It has been proposed that screening libraries for anti-TB targets should be less restricted and more diverse.[42]

2.2 Physicochemical properties of anti-mycobacterium natural products (949) reported in 1990-2012 timeframe

A dataset representing 814 compounds reported in the literature during the period from January 1990 to December 2011 as well as 135 compounds reported from January to December 2012 was compiled. Only compounds that exhibited MIC ≤ 64 μg ml^{-1} or ≥ 75% growth inhibition at 12.5 μg ml^{-1} or less[43] were chosen for physicochemical calculations. Histograms of the calculated physicochemical properties are depicted in Figure 2 (red bars). Out of 949 anti-mycobacterium natural products, 513 compounds showed no violation of Lipinski's parameters and the majority of the rest had violation in either log *P*, molecular weight, or both of those (Figure S2). The histogram of molecular weight (Figure 2a) followed a Gaussian distribution and showed a peak at 350-400 Da. About 30% of the analyzed compounds had molecular weights more than 500 Da. The molecular weight distribution was similar to that reported by Feher and Schmidt,[10] as well as to the molecular weight distribution of 126,140 natural products from *Dictionary of Natural Products* which was previously examined for Lipinski properties.[44]

Similar to the previous analysis of natural products,[10, 44] the calculated log *P* histogram of anti-mycobacterium natural products (Figure 2b) showed a normal distribution, however, a shift to a higher log *P* values was noticed for anti-mycobacterium natural products compared to all natural products. In contrast to TB drugs where 41% had negative calculated log *P*, a very small percentage (3%) of the active natural products showed a negative log *P* value. Another difference between these two sets was observed in the range of the log *P* -1 and 1. While about 33% of TB drugs have log *P* value between -1 and 1, only a small percentage (3.5%) of the anti-mycobacterium natural products possess these log *P* values. In the range of 0 to 3 the anti-mycobacterium natural products percentage increased gradually and reached a maximum 18% at log *P* 3 with the same percentage at log *P* 4 and 5 after which the distribution fell off. The histogram showed that around 34% of the compounds violated the Lipinski cutoff with a log *P* of over 5.

The distribution of hydrogen bond donors (HBD) is depicted in Figure 2c. Similar to TB drugs the distribution showed a peak at 1 followed by a steady decrease. About 6% of these anti-mycobacterium NPs had hydrogen bond donors over than 5. In contrast to TB drugs, the histogram of hydrogen bond acceptors (HBA) (Figure 2d) was Gaussian. It showed a peak at 3 that included 20% of the compounds. Only a small percentage of

anti-mycobacterium natural products (around 5%) had hydrogen bond donors higher than 5 and hydrogen bond acceptors greater than 10, while 18% of TB drugs were found to violate these cutoff numbers.

Polar surface area distribution (Fig. S3a) demonstrated a peak of distribution at 50-60 and about 12% of the analyzed compounds had a polar surface area of over 140. The distribution of the number of rotatable bonds (Fig. S3b) showed a wide range of flexibility. The maximum of distribution was in the range of 0 to 2 rotatable bonds with each possessing around 12-14% of the molecules. About 18% of the analyzed compounds exceed the number of rotatable bond cutoff of 10. The majority of those were molecules that possess aliphatic chains in their structure that makes the compounds more lipophilic.

While 77% of TB drugs obeyed Lipinski's Ro5, the proportion of anti-mycobacterium natural products was higher and about 85% of the compounds had no violation or just one violation of Lipinski's role of five (Figure 1). The result for anti-mycobacterium natural products is almost similar to the two previous analyses of natural products.[10, 44] In both investigations about 80% of analyzed compounds had less than two violations of Ro5.

3 Comparison of anti-mycobacterial natural product physicochemical space vs current anti-TB drugs – what does it tell us?

In order to compare the distribution of TB drugs and anti-mycobacterium natural products in physicochemical space, ChemGPS-NP[45] was used for principle component analysis (PCA) of the 949 anti-mycobacterium natural products and the 39 TB drugs (which included both the marketed and promising compounds in clinical trials). The first three principle components that explained 71% of the variance are plotted in Figures 3a and 3b. PC1 describes size, shape, and polarizability, PC2 represents aromatic and conjugation related properties and PC3 expresses lipophilicity, polarity and H-bond capacity.[45]

Figure 3a demonstrates the score plot of all TB drugs and shows that TB drugs occupy a very broad range of physicochemical space. The nature derived TB drugs however, can be divided into three clusters shown in cyan ovals. The first cluster is representative of polar and high molecular weight compounds comprised of aminoglycosides streptomycin, kanamycin, and amikacin and the cyclicpeptides capreomycin and viomycin. It is important to note that all of these compounds are injectable/intravenous drugs. Another cluster belongs to the low molecular weight compounds including cycloserine, pyrazinamide, isoniazide, and *p*-aminosalicylic acid. The molecular weight of these compounds does not exceed 200 Da. The last cluster of nature-based drugs contains rifamycines analogues, rifampicin and rifapentine. These orally bioavailable drugs possesses the highest molecular weight of TB drugs (MW of 822 and 877 Da respectively) with HBDs and HBAs exceeding the Lipinski's rule of five while having a log *P* value within the range.

The majority of TB drugs from synthetic sources fall into drug-like physicochemical space. Outliers of synthetic compounds are newly introduced compounds in which most of them have log *P* greater than 5 (TMC207, OPC-67683, clofazimine, and C215).

In order to evaluate the 949 anti-mycobacterium natural products reported in the 1990-2012 timeframe against TB drug physicochemical space, their first three PCs are plotted and compared to TB drugs in Figure 3b. Red dots show TB drugs and blue dots illustrate anti-mycobacterium natural products. Similar to the distribution of synthetic TB drugs, the majority of anti-mycobacterium NPs fall into drug-like space. Not many are toward the two major categories of nature-based TB drugs.

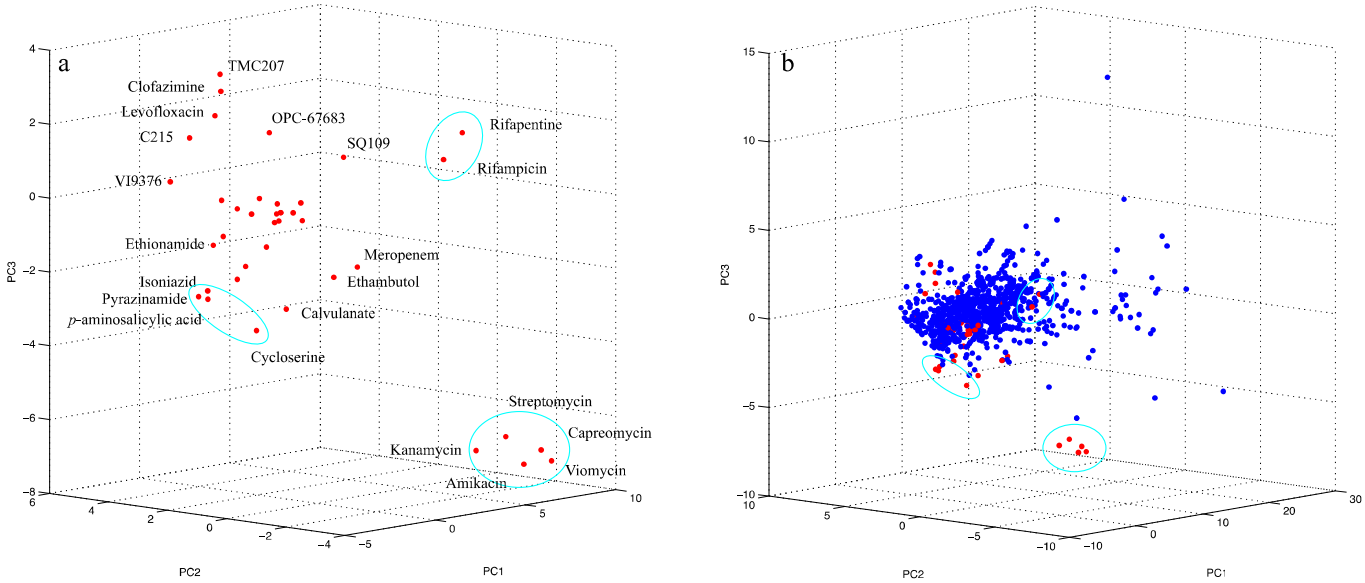


Figure 3 Score plot of TB drugs and anti-mycobacterium natural products. a) Physicochemical space of 39 approved TB drugs and candidates in clinical trials. Cyan ovals are showing the three clusters of natural product-based drugs while the rest of the compounds are synthetic TB drugs. b) Comparison of the physicochemical space of 949 anti-mycobacterium natural products (blue dots) with physicochemical space of 39 TB drugs (red dots).

Anti-mycobacterium natural products that occur close to the cluster of low molecular weight TB drugs are shown in Figure S4. The set comprises 23 compounds and the majority are phenolics. In addition there are five small molecular weight alkaloids containing a similar core ring of imidazolidine or pyrrolidine in their structures.

Log *P* plays an important role in drugability of anti-mycobacterium natural products. The mycobacterial cell wall contains a large amount of lipids mainly in form of mycolic acids that make a strong hydrophobic layer, which is believed to be responsible for the permeability barrier towards hydrophilic compounds.[46, 47]. The high percentage of anti-mycobacterium natural products (34%) with log *P* value more than 5, demonstrates the ability of lipophilic compounds to pass through the waxy cell wall of the mycobacteria in order to reach the target enzyme. We have shown that, in contrast to the TB drugs which display 33% of log *P* values between -1 and 1, only 3.5% of anti-mycobacterium natural products fall in this range. The set consists of 34 molecules that belong to different classes of compounds including phenolics and quinones, alkaloids, peptides and terpenes. The potent antibacterial thiazole peptides nocathiacins I-III (molecules **47-49** in supporting information) are among this set. Nocathiacins antibiotics are isolated from the fermentation broth of *Nocardia* sp. (ATCC202099)[48, 49] and a fungus *Amicolaptosis* sp. (MJ347-81F4).[50] These molecules exhibit strong activity towards a range of bacteria including *M. tuberculosis* (ATCC35828) with MIC value of ≤ 0.008 μg/ml.[51] The compounds however, have poor aqueous solubility.[51] Nocathiacin I was used as a lead for development of novel semi-synthetic water soluble nocathiacin analogs.[52-55]

4 Conclusions

Natural product-based TB drugs are vital components of first and second line TB drugs and have significant efficacy in TB treatment. However, the current contribution of nature-derived molecules in the TB drug pipeline is very small with only the repurposed nature-derived anti-TB compound rifapentine in clinical trials. Efforts of the past three decades for identification of anti-mycobacterium natural products have been very productive. About 949 secondary metabolites belonging to various chemical classes are reported to inhibit the growth of mycobacterium. Analysis of the physicochemical properties of these anti-mycobacterium natural products and comparison with the physicochemical properties of TB drugs revealed major differences between the log *P* distribution and the molecular weight distribution of these two sets. Anti-mycobacterium natural products showed a shift to higher log *P* values and only a small percentage (3%) had negative calculated log *P*, while a high percentage of TB drugs (41%) had negative log *P* values. TB drugs showed the maximum peak of distribution in the range of log *P* -1 to 1, while there were very few active NPs in this region. A small percentage of anti-mycobacterium natural products possess molecular weight less than 200 Da whereas many TB drugs have a molecular weight less than 200 Da. Principle component analysis of the 949 anti-mycobacterium natural products and the 39 TB drugs using ChemGPS-NP also highlighted that not many secondary metabolites are toward the cluster of the low molecular weight TB drugs. Of the 949 anti-mycobacterium natural products only 50 compounds cluster within the two clusters occupied by many TB drugs. The structures of the natural products in these two clusters, 23 and 34 compounds respectively, are given in the Supplementary Information. Within this set the most active compounds are MIC = 0.4 μg/ml ($H_{37}Ra$), MIC = 0.5 μg/ml (*M. smegmatis*), MIC = 8 μM ($H_{37}Rv$), [low MW] and MIC = 0.4 μg/ml ($H_{37}Ra$), MIC = 8.9 μM ($H_{37}Ra$), MIC = 1 μg/ml ($H_{37}Rv$), MIC = 17 μM ($H_{37}Rv$), MIC = 6.25 μg/ml (*M. tuberculosis*), MIC = 3.12 μg/ml (*M. vaccae*),), MIC = ≤ 0.008 μg/ml (*M. tuberculosis*) [log *P* between -1 and +1]. The drug space of existing TB drugs may help prioritisation of a continuing discovery pipeline of anti-mycobacterium natural product and hit-to-lead identification of anti-mycobacterium natural products reported in the literature.

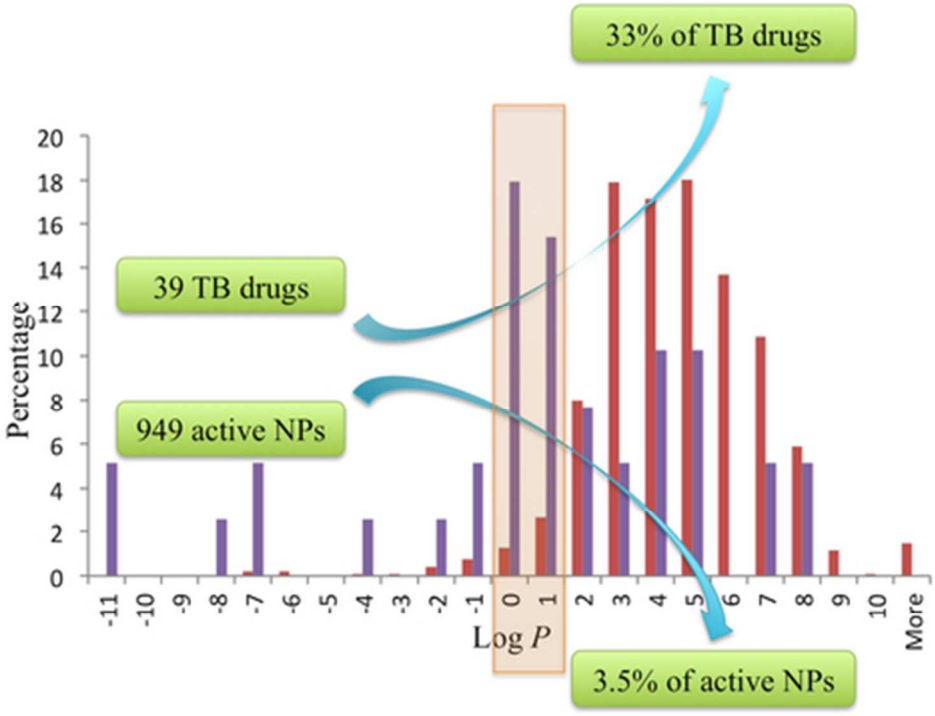
5 Acknowledgements

The work was supported by Bill & Melinda Gates Foundation Grand Challenges Explorations Grants OPP1008376 and OPP1035218 and the Australian Research Council (ARC) Linkage Project (LP120100485).

6 References

1. S. H. E. Kaufmann and P. van Helden, *Handbook of Tuberculosis: Clinics, Diagnostics, Therapy and Epidemiology*, Wiley, 2008.
2. W. H. Organization, *Treatment of Tuberculosis Guidelines*, 2007.
3. W. H. Organization, *The Global Plan to Stop TB, 2006-2015.*, 2006.
4. B. M. McArdle, M. R. Campitelli and R. J. Quinn, *J. Nat. Prod.*, 2006, **69**, 14-17.
5. B. M. McArdle and R. J. Quinn, *ChemBioChem*, 2007, **8**, 788-798.
6. E. Kellenberger, A. Hofmann and R. J. Quinn, *Nat. Prod. Rep.*, 2011, **28**, 1483-1492.
7. M. A. Koch and H. Waldmann, *Drug Discov. Today*, 2005, **10**, 471-483.
8. M. A. Koch, L. O. Wittenberg, S. Basu, D. A. Jeyaraj, E. Gourzoulidou, K. Reinecke, A. Odermatt and H. Waldmann, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 16721-16726.
9. R. Breinbauer, I. R. Vetter and H. Waldmann, *Angew. Chem. Int. Ed.*, 2002, **41**, 2879-2890.
10. M. Feher and J. M. Schmidt, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 218-227.
11. J. Hert, J. J. Irwin, C. Laggner, M. J. Keiser and B. K. Shoichet, *Nat. Chem. Biol.*, 2009, **5**, 479-483.
12. D. J. Newman and G. M. Cragg, *J. Nat. Prod.*, 2007, **70**, 461-477.
13. H. B. Woodruff and S. A. Waksman, *Ann. N. Y. Acad. Sci.*, 1960, **89**, 287-298.
14. S. A. Waksman, *Adv. Appl. Microbiol.*, 1969, **11**, 1-16.
15. M. J. Cron, O. B. Fardig, D. L. Johnson, H. Schmitz, D. F. Whitehead, I. R. Hooper and R. U. Lemieux, *J. Am. Chem. Soc.*, 1958, **80**, 2342-2342.
16. H. Kawaguchi, T. Naito, S. Nakagawa and K. I. Fujisawa, *J. Antibiot.*, 1972, **25**, 695-708.
17. W. J. Burman, K. Gallicano and C. Peloquin, *Clin. Pharmacokinet.*, 2001, **40**, 327-341.
18. F. A. Kuehl, F. J. Wolf, N. R. Trenner, R. L. Peck, R. P. Buhs, E. Howe, I. Putter, B. D. Hunnewell, R. Ormond, G. Downing, J. E. Lyons, E. Newstead, L. Chaiet and K. Folkers, *J. Am. Chem. Soc.*, 1955, **77**, 2344-2345.
19. J. Lehmann, *Lancet*, 1946, **1**, 15-16.
20. E. B. Herr, Jr. and M. O. Redstone, *Ann. N. Y. Acad. Sci.*, 1966, **135**, 940-946.
21. S. Nomoto, T. Teshima, T. Wakamiya and T. Shiba, *Tetrahedron*, 1978, **34**, 921-927.
22. B. W. Bycroft, *Chem. Commun.*, 1972, **0**, 660-661.

23. T. Kitagawa, T. Miura, K. Fujiwara and H. Taniyama, *Chem. Pharm. Bull.*, 1972, **20**, 2215-2225.
24. M. T. Gutierrez-Lugo and C. A. Bewley, *J. Med. Chem.*, 2008, **51**, 2606-2612.
25. A. Zumla, P. Nahid and S. T. Cole, *Nat. Rev. Drug Discov.*, 2013, **12**, 388-404.
26. N. Maggi, C. R. Pasqualucci, R. Ballotta and P. Sensi, *Chemotherapy*, 1966, **11**, 285-292.
27. W. H. Organization, *World Health Forum*, 1993.
28. W. H. Organization, *Global Tuberculosis Report* 2012.
29. B. R. Copp, *Nat. Prod. Rep.*, 2003, **20**, 535-557.
30. G. F. Pauli, R. J. Case, T. Inui, Y. Wang, S. Cho, N. H. Fischer and S. G. Franzblau, *Life Sci.*, 2005, **78**, 485-494.
31. A. L. Okunade, M. P. Elvin-Lewis and W. H. Lewis, *Phytochemistry*, 2004, **65**, 1017-1032.
32. B. R. Copp and A. N. Pearce, *Nat. Prod. Rep.*, 2007, **24**, 278-297.
33. A. García, V. Bocanegra-García, J. P. Palma-Nicolás and G. Rivera, *Eur. J. Med. Chem.*, 2012, **49**, 1-23.
34. C. E. Salomon and L. E. Schmidt, *Curr. Top. Med. Chem.*, 2012, **12**, 735-765.
35. C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Deliv. Rev.*, 1997, **23**, 3-25.
36. C. A. Lipinski, *Drug Discov. Today Technol.*, 2004, **1**, 337-341.
37. P. R. Andrews, D. J. Craik and J. L. Martin, *J. Med. Chem.*, 1984, **27**, 1648-1657.
38. G. Klebe and H. J. Böhm, *J. Recept. Sig. Transd.*, 1997, **17**, 459-473.
39. D. F. Veber, S. R. Johnson, H. Y. Cheng, B. R. Smith, K. W. Ward and K. D. Kopple, *J. Med. Chem.*, 2002, **45**, 2615-2623.
40. R. O'Shea and H. E. Moser, *J. Med. Chem.*, 2008, **51**, 2871-2878.
41. L. M. Fu and C. S. Fu-Liu, *Tuberculosis*, 2002, **82**, 85-90.
42. A. Koul, E. Arnoult, N. Lounis, J. Guillemont and K. Andries, *Nature*, 2011, **469**, 483-490.
43. J. Secrist, S. Anathan, C. Kwong, J. Maddry, R. Reynolds, A. Poffenberger, M. Michael, L. Miller, J. Krahenbuh, L. Adams, A. Biswas, S. Franzblau, D. Rouse, D. Winfield and J. Brooks, *Antimicrob. Agents Chemother.*, 2001, **45**, 1943-1946.
44. R. J. Quinn, A. R. Carroll, N. B. Pham, P. Baron, M. E. Palframan, L. Suraweera, G. K. Pierens and S. Muresan, *J. Nat. Prod.*, 2008, **71**, 464-468.
45. J. Larsson, J. Gottfries, S. Muresan and A. Backlund, *J. Nat. Prod.*, 2007, **70**, 789-794.
46. D. E. Minnikin, *Lipids: complex lipids, their chemistry, biosynthesis and roles*, Academic Press, London, 1982.
47. N. Rastogi, K. S. Goh and H. L. David, *Antimicrob. Agents Chemother.*, 1990, **34**, 759-764.
48. W. Li, J. E. Leet, A. H. A., D. R. Gustavson, D. M. Brown, L. Turner, K. Brown, J. Clark, H. Yang, J. Fung-Tomc and K. S. Lam, *J. Antibiot.*, 2003, **56**, 226-231.
49. J. E. Leet, W. Li, A. H. A., J. A. Maston, S. Huang, R. Huang, J. L. Cantone, D. Drexler, R. A. Dalterio and K. S. Lam, *J. Antibiot.*, 2003, **56**, 232-242.
50. T. Sasaki, T. Otani, H. Matsumoto, N. Unemi, M. Hamada, T. Takeuchi and M. Hori, *J. Antibiot.*, 1998, **51**, 715-721.
51. M. J. Pucci, J. J. Bronson, J. F. Barrett, K. L. DenBleyker, L. F. Discotto, J. C. Fung-Tomc and Y. Ueda, *Antimicrob. Agents Chemother.*, 2004, **48**, 3697-3701.
52. L. Xu, A. K. Farthing, J. F. Dropinski, P. T. Meinke, C. McCallum, E. Hickey and K. Liu, *Bioorg. Med. Chem. Lett.*, 2013, **23**, 366-369.
53. L. B. Xu, A. K. Farthing, J. F. Dropinski, P. T. Meinke, C. McCallum, P. S. Leavitt, E. J. Hickey, L. Colwell, J. Barrett and K. Liu, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 3531-3535.
54. A. Regueiro-Ren, B. N. Naidu, X. Zheng, T. W. Hudyma, T. P. Connolly, J. D. Matiskella, Y. Zhang, O. K. Kim, M. E. Sorenson, M. Pucci, J. Clark, J. J. Bronson and Y. Ueda, *Bioorg. Med. Chem. Lett.*, 2004, **14**, 171-175.
55. P. Hrnciar, Y. Ueda, S. Huang, J. E. Leet and J. J. Bronson, *J. Org. Chem.*, 2002, **67**, 8789-8793.



39x30mm (300 x 300 DPI)